

## **Ebola superspreading**

Christian L. Althaus, Ph.D.

Institute of Social and Preventive Medicine (ISPM), University of Bern,  
Finkenhubelweg 11, 3012 Bern, Switzerland

Email: [christian.althaus@alumni.ethz.ch](mailto:christian.althaus@alumni.ethz.ch); Phone: +41 31 631 56 40

Published in Lancet Infect Dis. 2015; 15(5):507-8.

DOI: [http://dx.doi.org/10.1016/S1473-3099\(15\)70135-0](http://dx.doi.org/10.1016/S1473-3099(15)70135-0)

Ousmane Faye and colleagues<sup>1</sup> recently described the chains of transmission for 152 individuals infected with Ebola virus diseases in Guinea. The resulting transmission trees provide unique insights into the individual variation in the number of secondary cases generated by an infected index case. A better understanding of this variation provides crucial information on epidemic spread, the expected number of superspreading events, and the effects of control measures.<sup>2</sup>

The number of secondary cases in the transmission trees is highly skewed, with 72% of individuals not generating further cases (figure). Fitting a negative binomial distribution to the data (appendix) provides maximum likelihood estimates of the mean (0.95, 95% CI 0.57-1.34) and the dispersion parameter ( $k = 0.18$ ; 95% CI 0.10-0.26). The mean corresponds to the basic reproduction number ( $R_0$ ) of the overall population. The estimated value of  $k$ , which is substantially smaller than 1, suggests that the distribution of the individual reproduction number is highly overdispersed.<sup>2</sup> The value for Ebola virus disease is similar to that estimated for severe acute respiratory syndrome ( $k = 0.16$ ).<sup>2</sup> This finding suggests that superspreading events for Ebola virus disease are an expected feature of the individual variation in infectiousness.<sup>3</sup>

I simulated stochastic trajectories of Ebola virus disease outbreaks starting from one infected index case (figure). To this end, I drew the number of secondary cases for each case from the fitted negative binomial distribution (appendix). The time from disease onset in one case to disease onset in the next case was drawn from the reported gamma-distributed serial interval with a mean duration of 15.3 days.<sup>4</sup> Although most outbreaks rapidly become extinct, some epidemic trajectories can reach to more than 100 infected cases. This finding is particularly remarkable because  $R_0$  is less than 1, and shows the potential for explosive outbreaks of Ebola virus disease.

$R_0$  during the early phase of the Ebola virus disease epidemic in Guinea has been estimated to be roughly 1.5.<sup>5</sup> The transmission trees from Faye and colleagues were generated from data obtained between February and August, 2014, when the reproduction number was fluctuating around unity.<sup>1,4</sup> That scenario is similar to the present situation in parts of west Africa where the incidence is declining but new

outbreaks still occur. The observed variation in individual infectiousness for Ebola virus disease means that although the probability of extinction is high, new index cases also have the potential for explosive regrowth of the epidemic.

### **Acknowledgements**

I received funding through an Ambizione grant from the Swiss National Science Foundation (project 136737).

### **Competing interests**

I declare no competing interests.

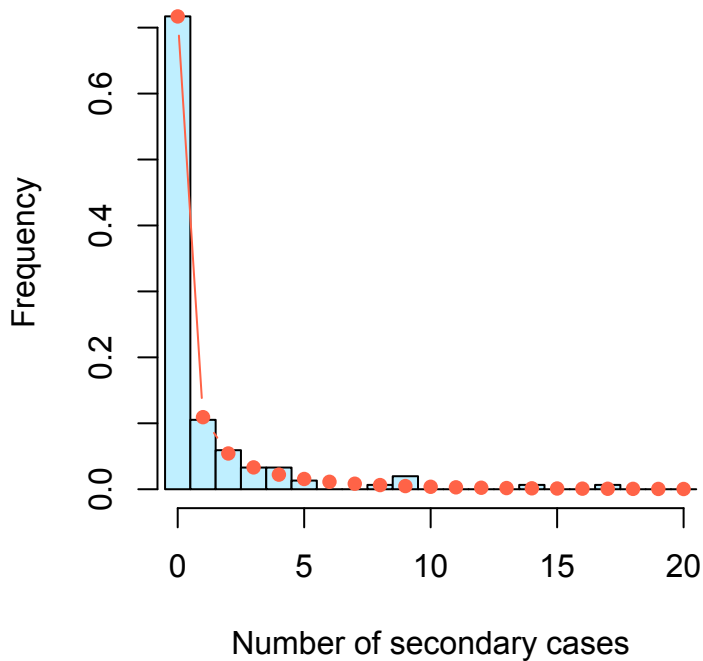
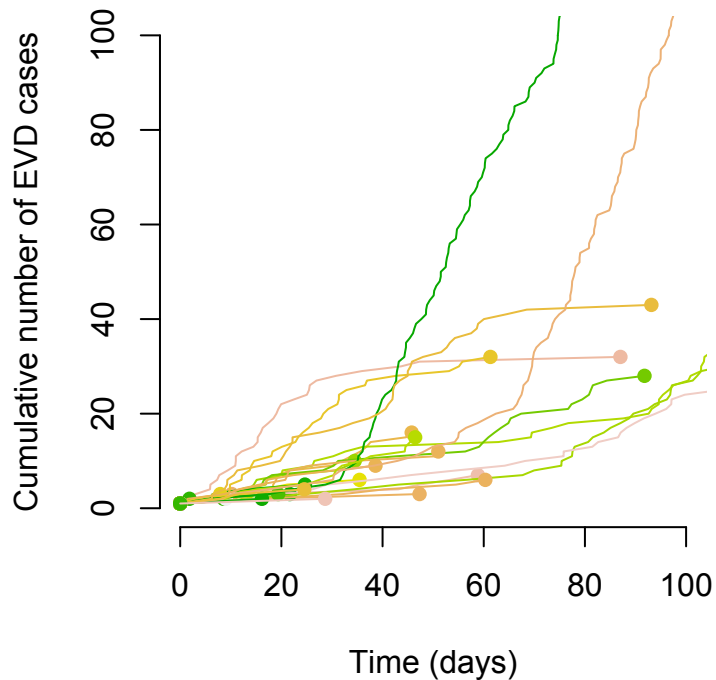
### **References**

- 1 Faye O, Boëlle P-Y, Heleze E, et al. Chains of transmission and control of Ebola virus disease in Conakry, Guinea, in 2014: an observational study. *Lancet Infect Dis* 2015; **15**: 320–26.
- 2 Lloyd-Smith JO, Schreiber SJ, Kopp ME, Getz WM. Superspreading and the effect of individual variation on disease emergence. *Nature* 2005; **438**: 355–59.
- 3 Volz EM, Pond SL. Phylodynamic analysis of Ebola virus in the 2014 Sierra Leone epidemic. *PLoS Curr* 2014; published online Oct 24. DOI:10.1371/currents.outbreaks.6f7025f1271 821d4c815385b08f5f80e.
- 4 WHO Ebola Response Team. Ebola virus disease in West Africa—the first 9 months of the epidemic and forward projections. *N Engl J Med* 2014; **317**: 1481–95.
- 5 Althaus CL. Estimating the reproduction number of Ebola virus (EBOV) during the 2014 outbreak in West Africa. *PLoS Curr* 2014; published online Sept 2. DOI:10.1371/ currents.outbreaks.91afb5e0f279e7f29e7056 095255b288.

## Figure legend

***Figure: Distribution of the number of secondary cases and outbreak trajectories for Ebola virus disease.***

(A) The histogram represents the observed frequencies in the number of secondary cases as given by the transmission trees in Faye and colleagues' study.<sup>1</sup> The line and dots correspond to the fitted negative binomial distribution. (B) Each line represents one of 200 stochastic realisations of epidemic trajectories. Dots show when the outbreak becomes extinct. A detailed analysis is reported in the appendix. EVD= Ebola virus disease.

**A****B**

# Appendix: Ebola superspreading

Christian L. Althaus ([christian.althaus@alumni.ethz.ch](mailto:christian.althaus@alumni.ethz.ch))

## Overview

This document describes the analysis of the number of secondary cases infected with Ebola virus disease (EVD), the serial interval distribution, and the resulting outbreak trajectories. All analyses were performed in the R software environment for statistical computing.<sup>1</sup>

```
library(fitdistrplus)
```

## Distribution of secondary cases

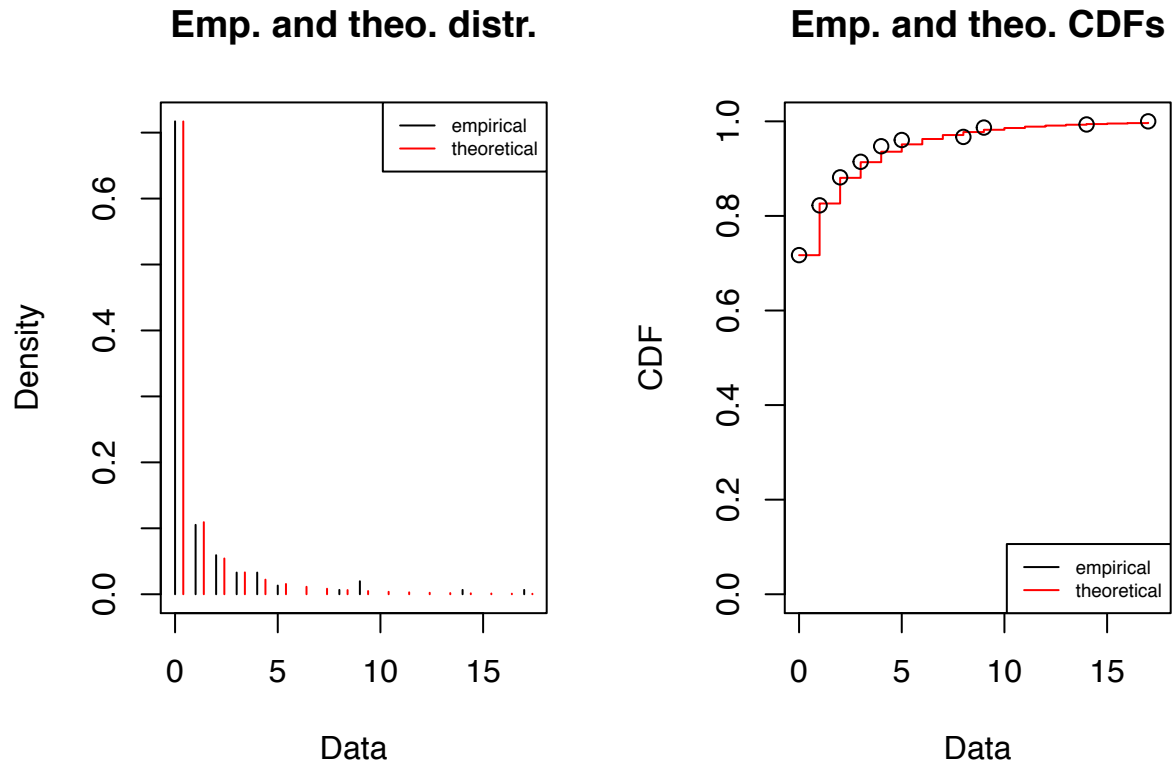
The transmission trees in figure 2D from Faye et al.<sup>2</sup> can be used to obtain the distribution of the number of secondary cases generated by an infected index case.

```
# Number of individuals in the trees
n <- 152
# Number of secondary cases for all individuals
c1 <- c(1,2,2,5,14,1,4,4,1,3,3,8,2,1,1,4,9,9,1,1,17,
        2,1,1,1,4,3,3,4,2,5,1,2,2,1,9,1,3,1,2,1,1,2)
c0 <- c(c1,rep(0,n-length(c1)))
# Fitting a negative binomial distribution to the number of secondary cases
fit.cases <- fitdist(c0,"nbinom")
summary(fit.cases)
```

```
## Fitting of the distribution ' nbinom ' by maximum likelihood
## Parameters :
##      estimate Std. Error
## size    0.1814      0.0399
## mu      0.9538      0.1981
## Loglikelihood: -177.2    AIC: 358.4    BIC: 364.4
## Correlation matrix:
##           size      mu
## size 1.0000000 0.0001384
## mu    0.0001384 1.0000000
```

Fitting a negative binomial distribution to the number of secondary cases provides maximum likelihood estimates of the mean (0.95, 95% confidence interval [CI]: 0.57-1.34) and the dispersion parameter  $k = 0.18$  (95% CI: 0.1-0.26).

```
plot(fit.cases)
```



**Figure 1:** Empirical and theoretical density distribution and the cumulative density function (CDF) of the number of secondary cases.

## Serial interval distribution

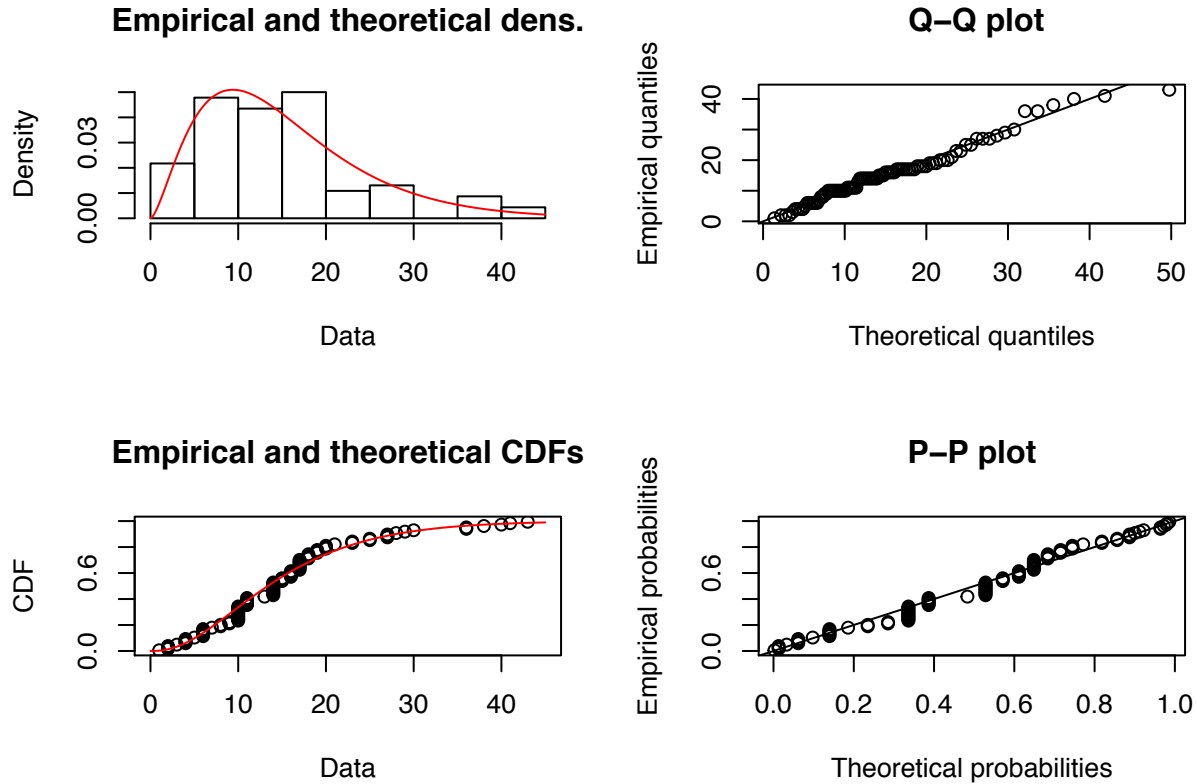
The serial interval distribution of EVD can be obtained from figure 3E of the study by the WHO Ebola Response Team.<sup>3</sup>

```
# Range of reported serial intervals
days <- 0:43
# Observed intervals for each day
frequency <- c(0,1,3,1,4,1,6,1,2,2,11,6,0,1,10,3,5,8,4,3,3,1,
               0,2,0,2,0,3,1,1,1,0,0,0,0,0,2,0,1,0,1,1,0,1)
d <- rep(days,frequency)
# Fitting a gamma distribution to the serial interval
fit.serial <- fitdist(d,"gamma")
summary(fit.serial)

## Fitting of the distribution ' gamma ' by maximum likelihood
## Parameters :
##      estimate Std. Error
## shape  2.5931    0.36042
## rate   0.1697    0.02602
## Loglikelihood: -324.6   AIC:  653.1   BIC:  658.2
## Correlation matrix:
##      shape  rate
## shape 1.0000 0.9065
## rate  0.9065 1.0000
```

Fitting a gamma distribution to the data provides maximum likelihood estimates of the mean serial interval (15.28 days) and the shape parameter (2.59).

```
plot(fit.serial)
```



**Figure 2:** Empirical and theoretical density distribution and the cumulative density function (CDF) of the serial interval.

## Simulating outbreaks

Based on the serial interval distribution and the number of secondary cases, one can simulate stochastic trajectories of EVD outbreaks starting from a single infected index case.

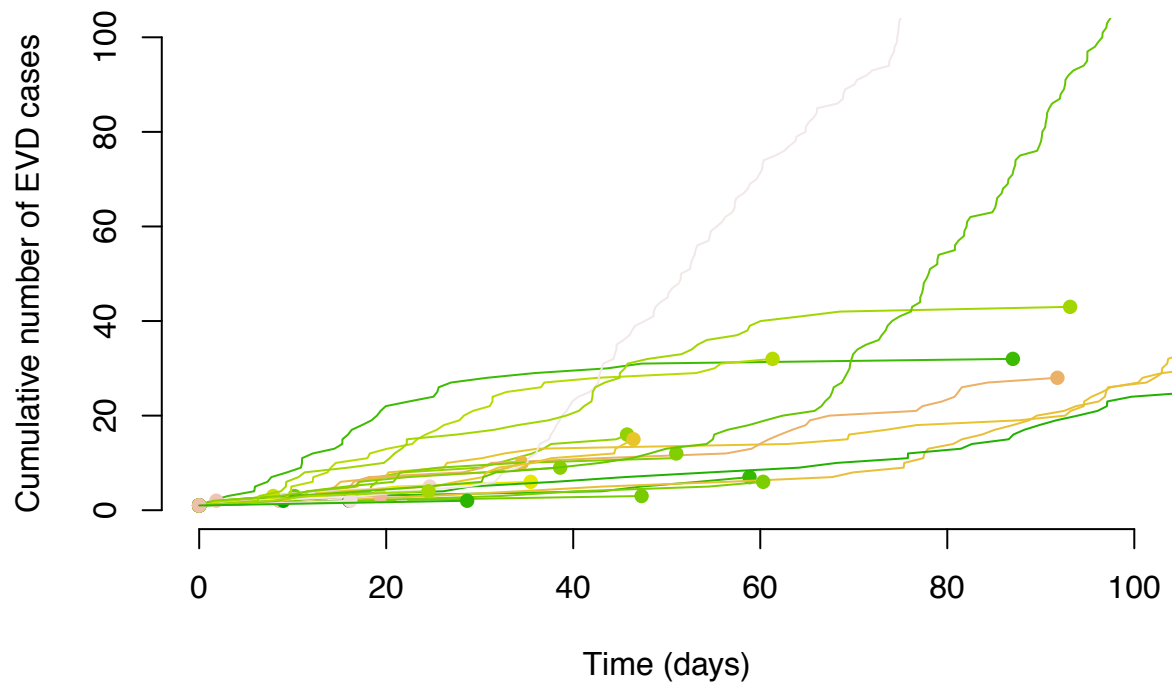
```
# Set seed for random number generator
set.seed(645)
# Number of simulation runs
runs <- 1e2
# Number of initial cases
seed <- 1
# Initialize plot
plot(NA,xlim=c(0,100),ylim=c(0,100),xlab="Time (days)",
     ylab="Cumulative number of EVD cases",frame=FALSE)
# Set color scheme for different trajectories
cols <- sample(terrain.colors(runs))
# Simulate outbreak trajectories
for(i in 1:runs) {
  cases <- seed
```



```

t <- rep(0,seed)
times <- t
while(cases > 0) {
  secondary <- rnbinom(cases,size=fit.cases$estimate[1],mu=fit.cases$estimate[2])
  t.new <- numeric()
  for(j in 1:length(secondary)) {
    t.new <- c(t.new,t[j] + rgamma(secondary[j],shape=fit.serial$estimate[1],
    rate=fit.serial$estimate[2]))
  }
  cases <- length(t.new)
  t <- t.new
  times <- c(times,t.new)
}
lines(sort(times),1:length(times),col=cols[i],lwd=1)
points(max(times),length(times),col=cols[i],pch=16)
}

```



**Figure 3:** Simulated outbreaks of EVD. Each line represents one of 200 stochastic realizations of epidemic trajectories. Dots indicate that the outbreak goes extinct.

## References

- 1R Development Core Team *et al.* R: A Language and Environment for Statistical Computing. Vienna, Austria: R Foundation for Statistical Computing, 2014.
- 2Faye O, Boëlle P-Y, Heleze E *et al.* Chains of transmission and control of Ebola virus disease in Conakry, Guinea, in 2014: an observational study. *Lancet Infect Dis* 2015; published online Jan. DOI:[10.1016/S1473-3099\(14\)71075-8](https://doi.org/10.1016/S1473-3099(14)71075-8).
- 3WHO Ebola Response Team *et al.* Ebola virus disease in West Africa—the first 9 months of the epidemic and forward projections. *N Engl J Med* 2014; **371**: 1481–95.